

RESEARCH ARTICLE

# Fully Automated Enhanced Tumor Compartmentalization: Man vs. Machine Reloaded

Nicole Porz<sup>1,2</sup>\*, Simon Habegger<sup>1</sup>\*, Raphael Meier<sup>3</sup>, Rajeev Verma<sup>1</sup>, Astrid Jilch<sup>2</sup>, Jens Fichtner<sup>2</sup>, Urspeter Knecht<sup>1</sup>, Christian Radina<sup>2</sup>, Philippe Schucht<sup>2</sup>, Jürgen Beck<sup>2</sup>, Andreas Raabe<sup>2</sup>, Johannes Slotboom<sup>1</sup>, Mauricio Reyes<sup>3</sup>‡, Roland Wiest<sup>1</sup>‡

**1** Support Center for Advanced Neuroimaging—Institute for Diagnostic and Interventional Neuroradiology, University Hospital Inselspital and University of Bern, Bern, Switzerland, **2** Department of Neurosurgery, University Hospital Inselspital and University of Bern, Bern, Switzerland, **3** Institute of Surgical Technology and Biomechanics, University of Bern, Bern, Switzerland

\* These authors contributed equally to this work.

‡ These authors share senior authorship on this work.

\* [simon.habegger@insel.ch](mailto:simon.habegger@insel.ch)



## OPEN ACCESS

**Citation:** Porz N, Habegger S, Meier R, Verma R, Jilch A, Fichtner J, et al. (2016) Fully Automated Enhanced Tumor Compartmentalization: Man vs. Machine Reloaded. PLoS ONE 11(11): e0165302. doi:10.1371/journal.pone.0165302

**Editor:** Han-Chiao Isaac Chen, University of Pennsylvania, UNITED STATES

**Received:** April 12, 2016

**Accepted:** October 10, 2016

**Published:** November 2, 2016

**Copyright:** © 2016 Porz et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper.

**Funding:** This study was supported by the Swiss National Science Foundation (SNF Grant No. 320030\_140958), the Bernese Cancer League, the Swiss Cancer League and the European Unions Seventh Framework Programme for research, technological development and demonstration under grant agreement No. 600841. Brainlab AG provided support in the form of salary for author CR, but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The

## Abstract

### Objective

Comparison of a fully-automated segmentation method that uses compartmental volume information to a semi-automatic user-guided and FDA-approved segmentation technique.

### Methods

Nineteen patients with a recently diagnosed and histologically confirmed glioblastoma (GBM) were included and MR images were acquired with a 1.5 T MR scanner. Manual segmentation for volumetric analyses was performed using the open source software 3D Slicer version 4.2.2.3 ([www.slicer.org](http://www.slicer.org)). Semi-automatic segmentation was done by four independent neurosurgeons and neuroradiologists using the computer-assisted segmentation tool SmartBrush® (referred to as SB), a semi-automatic user-guided and FDA-approved tumor-outlining program that uses contour expansion. Fully automatic segmentations were performed with the Brain Tumor Image Analysis (BraTumIA, referred to as BT) software. We compared manual (ground truth, referred to as GT), computer-assisted (SB) and fully-automated (BT) segmentations with regard to: (1) products of two maximum diameters for 2D measurements, (2) the Dice coefficient, (3) the positive predictive value, (4) the sensitivity and (5) the volume error.

### Results

Segmentations by the four expert raters resulted in a mean Dice coefficient between 0.72 and 0.77 using SB. BT achieved a mean Dice coefficient of 0.68. Significant differences were found for intermodal (BT vs. SB) and for intramodal (four SB expert raters) performances. The BT and SB segmentations of the contrast-enhancing volumes achieved a

specific roles of these authors are articulated in the 'author contributions' section.

**Competing Interests:** CR is working as Clinical Consultant for Brainlab Sales GmbH Germany. The SmartBrush® software evaluated in this study is commercialized by Brainlab AG. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

high correlation with the GT. Pearson correlation was 0.8 for BT; however, there were a few discrepancies between raters (BT and SB 1 only). Additional non-enhancing tumor tissue extending the SB volumes was found with BT in 16/19 cases. The clinically motivated sum of products of diameters measure (SPD) revealed neither significant intermodal nor intra-modal variations. The analysis time for the four expert raters was faster (1 minute and 47 seconds to 3 minutes and 39 seconds) than with BT (5 minutes).

## Conclusion

BT and SB provide comparable segmentation results in a clinical setting. SB provided similar SPD measures to BT and GT, but differed in the volume analysis in one of the four clinical raters. A major strength of BT may its independence from human interactions, it can thus be employed to handle large datasets and to associate tumor volumes with clinical and/or molecular datasets ("-omics") as well as for clinical analyses of brain tumor compartment volumes as baseline outcome parameters. Due to its multi-compartment segmentation it may provide information about GBM subcompartment compositions that may be subjected to clinical studies to investigate the delineation of the target volumes for adjuvant therapies in the future.

## Introduction

Volumetry of malignant brain tumors (glioblastoma multiforme (GBM)) is usually performed semi-automatically or by manual delineation by an expert in a clinical setting. The latter option is hampered by time and personnel costs, as well as by inter-rater variability [1]. Volumetry is frequently integrated into treatment planning (e.g. to inform the neurosurgeon about tumor location for operation planning, and radiation oncologists to support therapy planning [2–4]). GBM subcompartment volume analysis may aid in biophysical modeling of brain tumor infiltration [5]. Appropriate assessment of the extent of resection plays a role in the prognosis of GBMs, since maximizing the extent of resection influences survival in glioblastoma patients. A complete resection of enhancing tumor, defined as the removal of the final 1–2% of the tumor, seems to provide the most benefit in terms of survival [6–8]. Further, there is increasing evidence that tumor expansion beyond areas of blood–brain barrier disruption (i.e. the enhancing compartment) impacts survival of patients with a GBM [9–11] and should be considered in pre-surgical planning [12].

Fully-automated user-independent segmentation tools are currently employed predominantly for research [13]. Semi-manual image-guided contouring software is routinely used in many operating theaters and in radiation oncology, but still requires user-dependent manual interaction.

We recently devised the fully-automated multimodal segmentation tool BratTumIA (BT) for pre-surgical and longitudinal tumor segmentation [14, 15]. BT calculates tumor volumes of healthy and tumor tissue compartmentalization with a variability that is comparable to that achieved with time-consuming manual tumor delineation by an expert [13, 14]. In this study, we aimed to investigate the performance of fully- vs. semi-automatic brain tumor volumetry. For comparison, we employed SmartBrush® (SB), which is routinely used for surgical planning and volumetric quantification of pre- and postoperative tumor volumes [9, 16]. We determined whether: (i) 2D diameter-based and (ii) 3D volumetric-based criteria for brain tumor assessment can be reliably calculated with BT and SB, (iii) whether the (semi-)automatic segmentations are

comparable with an expert-rater-based ground truth (GT) in terms of the Dice coefficient, and (iv) whether the subcompartment analysis by BT, which takes into account the infiltrative growth patterns of GBMs, improves the routine estimations of gross tumor volume.

## Materials and Methods

### Study population

Patients with a newly diagnosed and histologically confirmed GBM, enrolled preoperatively at our institution between October 2012 and July 2013, underwent prospective manual segmentation, semi-automatic and automated volumetry. The manual GT was previously used in part by Porz et al. [15].

Exclusion criteria were: incomplete image acquisition, Karnofski performance status < 70%, abnormal hematologic, renal or hepatic function, and previous cranial neurosurgery. The study was approved by the Local Research Ethics Commission “Kantonale Ethik-kommission Bern”. All patients provided written informed consent.

### MR imaging protocol

MR images were acquired with similar protocols and field strengths on two 1.5 T MR scanners from one vendor (Siemens Avanto and Siemens Aera, Siemens, Erlangen, Germany). Every patient was subject to a standardized MR imaging protocol including: (1) pre-contrast 3D-T1w-multiplanar reconstruction (MPR) in sagittal acquisition, 1 mm isotropic resolution; (2) post-contrast 3D-T1w-MPR in sagittal acquisition, 1 mm isotropic resolution; (3) 3D-T2w (SPC) in sagittal acquisition, 1 mm isotropic resolution; and (4) fluid-attenuated inversion recovery (FLAIR) (2D turbo inversion recovery) in axial acquisition. The sequence parameters were: (1) for pre-contrast 3D-T1w MPR sequences echo time (TE) = 2.67 ms, repetition time (TR) = 1580 ms, field of view (FOV) =  $256 \times 256 \text{ mm}^2$ , flip angle (FA) =  $8^\circ$ , with an isotropic voxel resolution of  $1 \times 1 \times 1 \text{ mm}$ ; (2) for post-contrast T1w TE = 4.57 ms, TR = 2070 ms, FOV =  $256 \times 256 \text{ mm}^2$ , FA =  $15^\circ$ , using isotropic  $1 \times 1 \times 1 \text{ mm}$  voxels; (3) for 3D-T2w (SPC) in sagittal acquisition TE = 380 ms, TR = 3000 ms, FOV =  $256 \times 256 \text{ mm}^2$ , FA =  $120^\circ$ , using isotropic  $1 \times 1 \times 1 \text{ mm}$  voxels; (4) for 2D FLAIR sequence TE = 80 ms, TR = 8000 ms, FOV =  $256 \times 256 \text{ mm}^2$ , FA =  $120^\circ$ , using a non-isotropic voxel size of  $1 \times 1 \times 3 \text{ mm}$ .

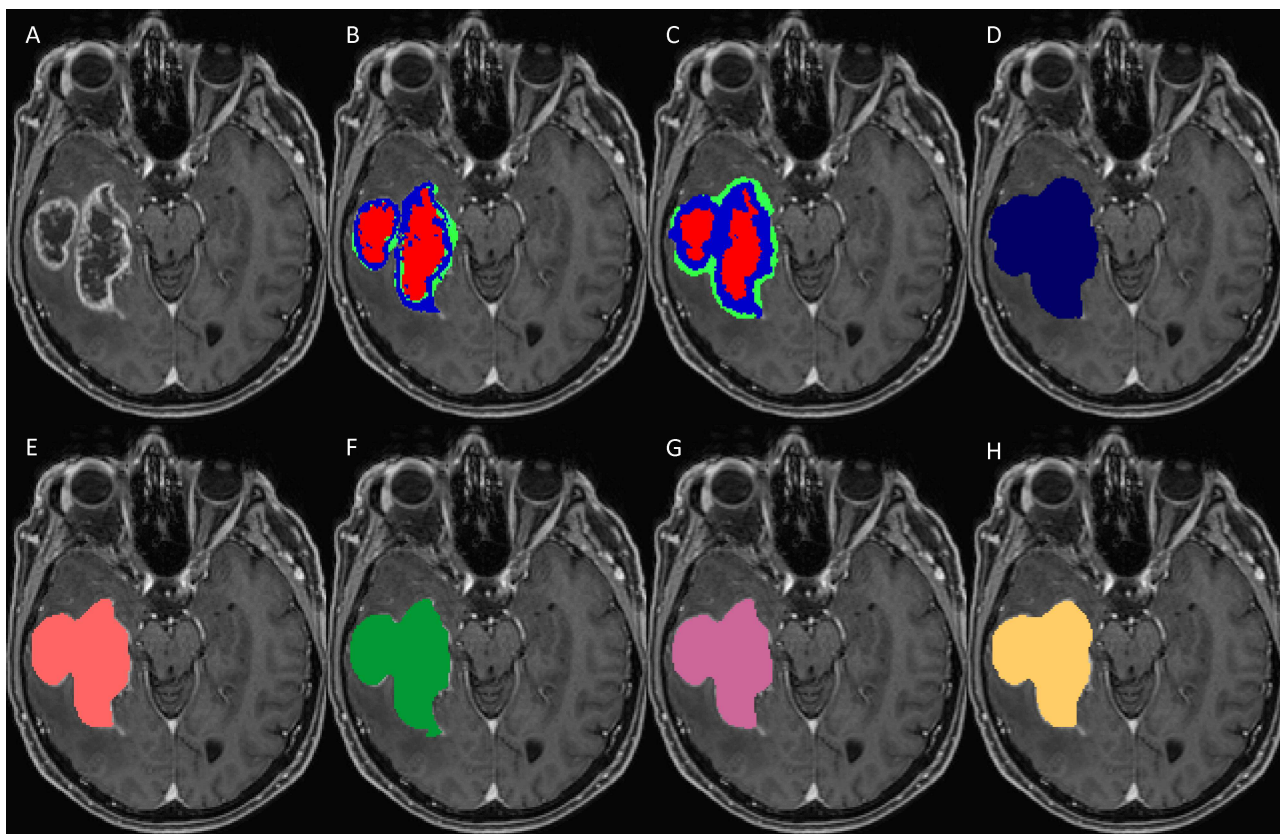
### Comparative metrics

Different metrics were employed to assess the segmentation quality of the two methods and of the four SB raters. The quantitative measures computed included Dice coefficient [17], absolute and relative volume, sum of products of squared diameters (SPD), sensitivity and positive predictive value (PPV). The Dice coefficient is a value between zero and one that expresses the amount of overlap between two segmentations, with one being a perfect overlap. It can be considered a standard metric in image analysis [17]. Besides the volumes of the individual segmentations, the relative volume, which is the segmented volume subtracted from the GT volume, was also evaluated. The SPD metric was mainly considered due to its prevalence in clinical practice, which in turn is due to its capability to provide a fast assessment of the tumor growth rate. Moreover, the World Health Organization recommends use of the SPD for gross tumor volume [18] and refined response assessment [19]. The segmentation sensitivity states what proportion of actual tumor was detected by the method and/or rater [20]. Last but not least, the PPV denotes the correctly segmented tumor divided by segmented tumor [21]. Sensitivity and PPV again yield values in the range between zero and one. Values closer to one indicate a better performance.

Although the objective of this study was to compare a fully-automatic brain tumor segmentation method with a commercially available semi-automatic one, it was decided not to merge the data from the four SB raters. Segmentation variability is always an issue one has to consider when dealing with methods that require human interaction and should therefore not be neglected in a comparative study. Furthermore, the decision was taken on the basis that no merging method should be applied unless out of necessity with respect to the main objective of the study, as such merging is always subject to data loss. Finally, the decision is supported by the results obtained during this study and by [22] where issues with the widely used simultaneous truth and performance level estimation (STAPLE) merging algorithm are discussed.

## Manual volumetry

Manual segmentation was performed by a neurosurgeon experienced in brain tumor analysis and supervised by a neuroradiologist with more than 15 years of experience in brain tumor imaging. The neurosurgeon employed the open source software 3D Slicer Version 4.2.2.3 ([www.slicer.org](http://www.slicer.org)) [23]. The images from the 19 patients were segmented manually slice by slice (Fig 1(B)). Segmentation was performed on T1w, T1wGd (Fig 1(A)), T2w and FLAIR sequences according to the VASARI MR feature guide v.1.1 (<https://wiki.nci.nih.gov/display/CIP/VASARI>). For intermodal comparisons and analysis, the gross tumor volume (TV)–encompassing the enhancing part, the non-enhancing part and the necrotic core of the GBM–



**Fig 1. Set of MRI sequences used in this study for manual, automatic, and semi-automatic tumor volumetry.** Original T1-weighted post-contrast MRI slice (A), manual subcompartmental segmentation into non-enhancing tumor (green), enhancing tumor (blue), and necrotic tissue (red) (B). BT subcompartmental segmentation into non-enhancing tumor (green), enhancing tumor (blue) and necrotic tissue (red) (C). BT core tumor segmentation (dark blue, D), SB1 core tumor segmentation (light red, E), SB2 core tumor segmentation (green, F), SB3 core tumor segmentation (purple, G) and SB4 core tumor segmentation (yellow, H).

doi:10.1371/journal.pone.0165302.g001

was selected. Manual segmentation was defined as the GT for further analyses [13]. Note that in the rest of this article we use GT and manual segmentations interchangeably.

## Automated segmentation

The fully automatic segmentations of this study were performed using BT (<https://www.nitrc.org/projects/bratumia/>). BT integrates a pipeline of three distinct parts, namely the preprocessing, classification and regularization units. The software takes as its input the widely acquired structural T1, T1-contrast (T1w), T2 and FLAIR MRI sequences. The preprocessing unit aligns these sequences to a common position and resolution through registration [24]. Also, a brain extraction mask computed from the T1w volume is applied to the four sequences. The output is forwarded to the classification unit. For every voxel in each volume a number of features, such as the mean intensity in its neighborhood, are computed. A random forest classifier then decides, based on the features exhibited by a voxel, which tissue type it depicts. The last unit is necessary to reduce implausible or even impossible tissue constellations between neighboring voxels. The core of BT evolved from [25] and the methodology was previously described in [14]. BT is able to distinguish seven brain tissue types, three healthy (gray matter, white matter and cerebrospinal fluid) and four tumor (edema, necrosis, non-enhancing tumor and enhancing tumor) tissues (Fig 1C and 1D; healthy tissues are not shown). Along the segmentations, BT further outputs the skull-stripped structural sequences and a report file with, among other information, the individual tissue volumes and the SPD value.

To illustrate the functionality and potential application of BT, we present the example of its additional use during biopsy in a GBM case. Biopsy was performed with the frameless neuro-navigation system (Brainlab® VarioGuide). BT segmentations were loaded to the iPlan (3.0.2 cranial) software as an overlay to the post-contrast T1 MRI sequence. The results of the sub-compartment analysis (i.e. of the enhancing part, the non-enhancing T2/FLAIR-hypointense part, the T2/FLAIR hyperintense vasogenic edema and the necrotic core) were stored in Brainlab® native object format. During surgery, it is possible to add the subcompartment segmentation objects as an overlay to the original MRI sequence, according to the surgeon's requirements (see Fig 2).

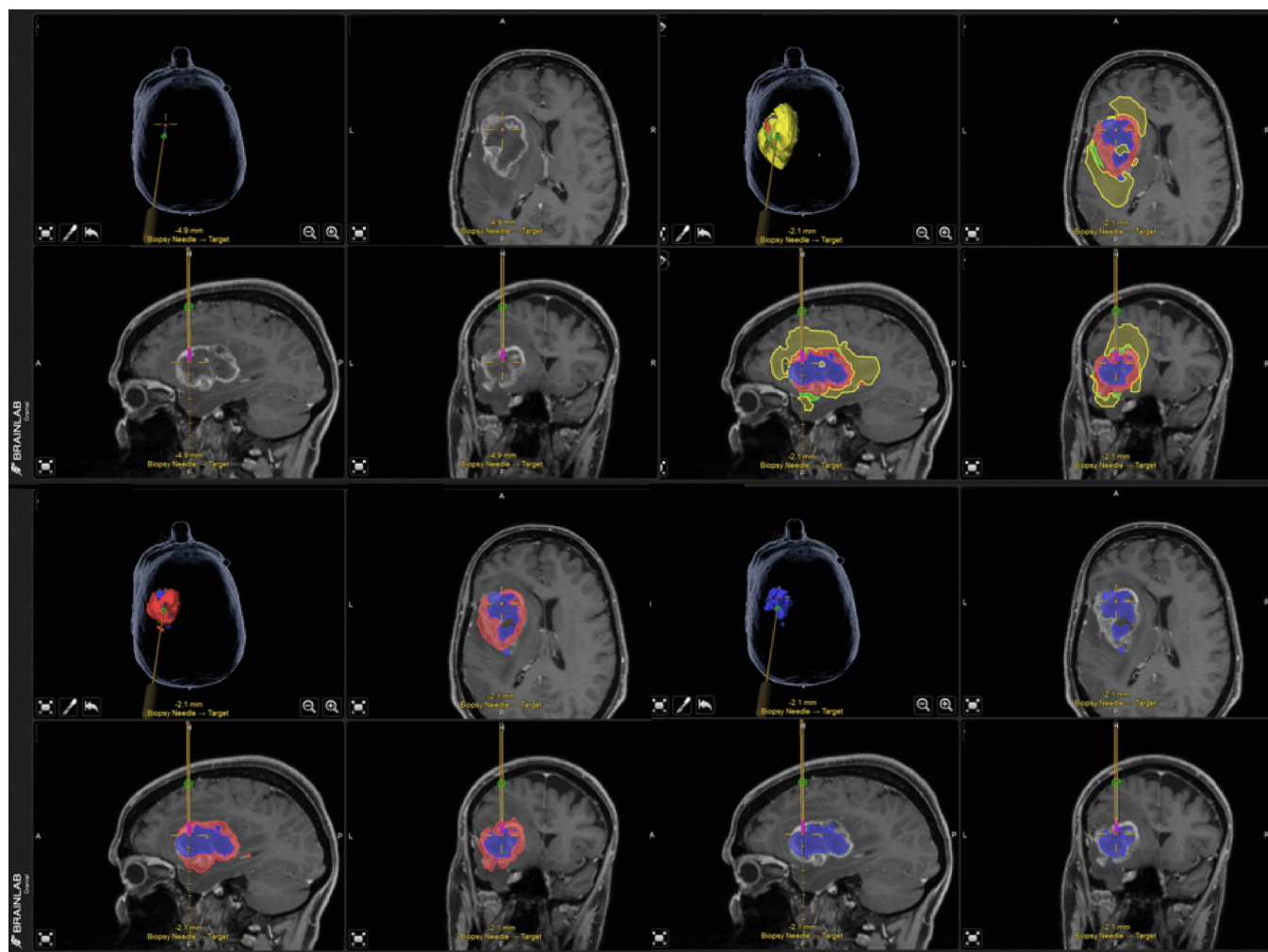
## Manual segmentation using SB

The semi-automatic segmentation was performed by four expert raters, three neurosurgeons (SB1, SB2 and SB4) and one neuroradiologist (SB3) (Fig 1E and 1H). All the neurosurgeons had more than five years of experience in semi-automatic brain tumor analysis and the neuroradiologist had more than five years of experience in MR reading of gliomas. All manual segmenters were instructed by an expert neuroradiologist with more than 15 years of experience in brain tumor imaging. The raters employed the contour-expansion-based, semi-automatic SB. It utilizes an intelligent region growing algorithm which, guided by the user, extends the brushed area to neighboring areas with similar intensities. The tumor has to be segmented in this manner on at least two slices in two different views. This provides sufficient input for the tool to compute the corresponding 3D tumor volume. The result can be further improved by extending or erasing the computed segmentation on individual slices.

## Statistical methods

For a first overview with a one-way analysis of variance, the non-parametric Kruskal-Wallis test was applied due to the non-normally distributed character of the data. The pairwise significance test was performed using the two-sided Wilcoxon signed-rank test [26]. The p-values were corrected for multiple comparisons with the Bonferroni method [27]. The chosen





**Fig 2. Stereotactic biopsy with the frameless neuronavigation system (Brainlab® VarioGuide) using BraTumIA segmentation.** The figures indicate the T1w raw image and the BT subcompartment overlays during biopsy. Upper row left column: original T1wGd without tumor delineation, right column: all automatic segmented tumor subcompartments are visible. Bottom row left column: necrosis and contrast-enhancing tumor volume, right column: only necrotic tumor volume. Color code for segmentations: red = enhancing tumor, yellow = edema, blue = necrosis and green = non-enhancing tumor.

doi:10.1371/journal.pone.0165302.g002

significance level was  $\alpha = 0.05\%$ . The interrelationship between the manual, semi- and fully-automatic segmentations was determined with the Pearson correlation coefficient. The statistical analysis was carried out with RStudio (<http://www.rstudio.com>).

## Results

### Study population

The mean age of patients at preoperative MR imaging was 65 years (range 37–76 years) and the mean pre-operative Karnofski performance status was  $85 \pm 26\%$  (range 70–90%). Nine of the 19 patients were female. Five patients underwent stereotactic biopsy, six subtotal resections and eight complete resections of enhancing tumor. All diagnoses were confirmed by histopathology.

## SPD

The differences between the SB/BT segmentations and the GT are depicted in Fig 3. BT achieved a smaller interquartile range (IQR) than the four SB raters did. The median difference from the GT was lower for the SB segmentations. The rather large difference between median and mean (red asterisk) for SB3 and BT hints at a skewed distribution of the data (SPD differences) and/or strong outliers. In contrast, the differences in SPD for SB1, SB2 and SB4 appeared to be evenly distributed. No rater or method introduced strong outliers indicating overestimation, whereas SB2 and BT included underestimation outliers. Only SB2 showed a tendency to underestimate the SPD value. SB1 showed, in general, no such tendency. SB3 and SB4 as well as BT were more inclined to overestimate the 2D SPD measure.

Differences between the expert raters and methods were not significant (Fig 4).

## Dice coefficient

The analysis of the Dice coefficient indicated a mean GT overlap of the four SB raters between 0.72 and 0.77 and of 0.68 for BT. An overview of these results is provided in Fig 5. The SB4 segmented similarly to BT. The outliers were similar for the SB1, SB2 and SB3, but differed from those of SB4 and BT.

The statistical difference or similarity of the raters/methods in terms of Dice coefficient is depicted in Fig 4(A). One can observe that there was a significant inter-rater variability during semi-automatic segmentation. The qualitative observation that rater SB4 was closest to BT was quantitatively confirmed with the result shown in Fig 4(A) (no connection between BT and SB4).

## Volume

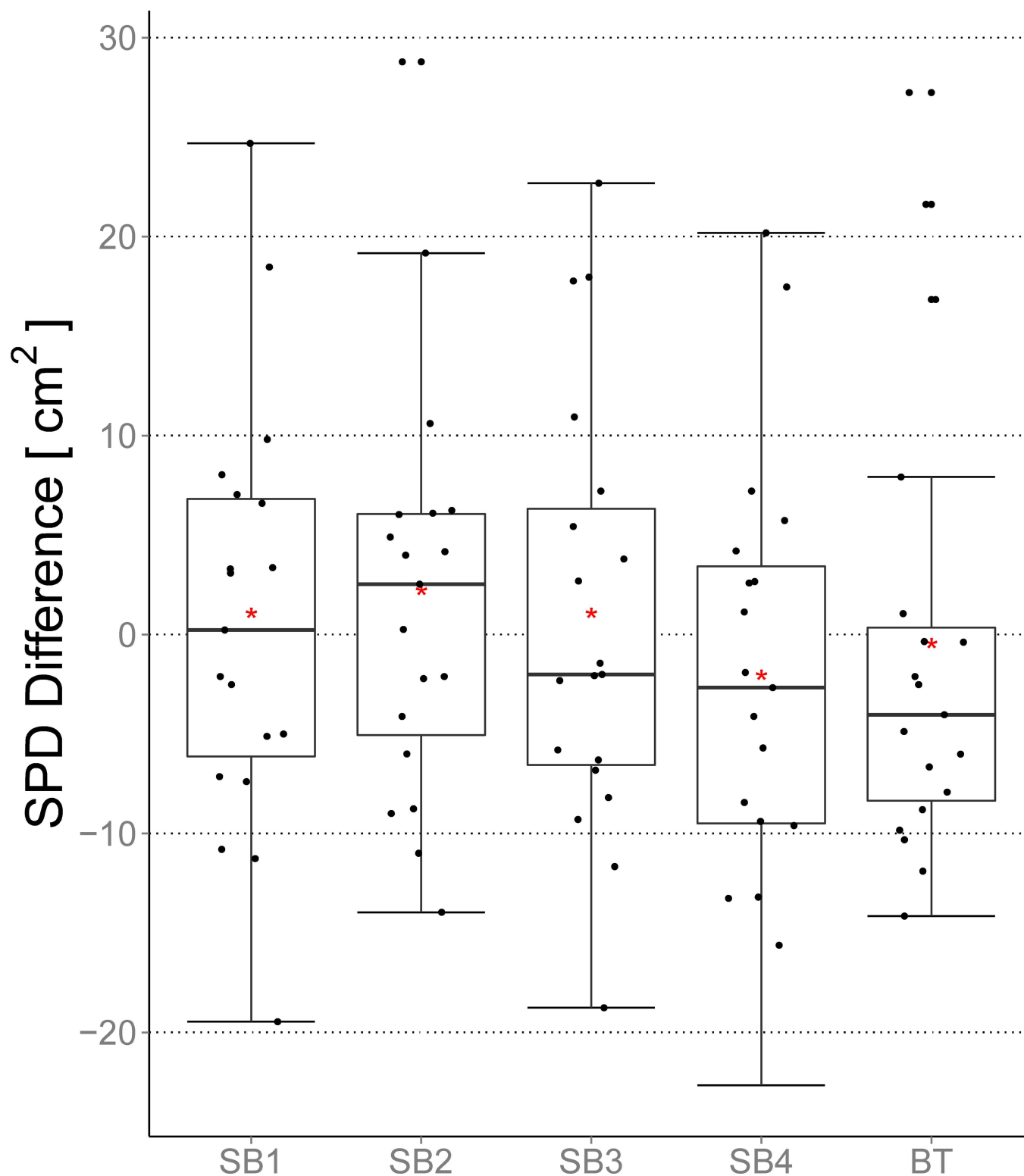
The volume differences to GT are illustrated in Fig 5. For BT we identified no systematic bias toward over- or underestimation of the volume (mean/median close to zero). SB raters tended to underestimate the volumes (see Fig 6). The volumes were tested for significant differences between all ten possible combinations of raters and methods. The result is depicted in Fig 4(B). A weakly significant difference (p-value rather close to  $\alpha = 0.05\%$ ) could be found only for BT vs. SB2. As a next step, we analyzed the volumes of additional non-enhancing tumor tissue as provided by the subcompartment analysis of BT compared to the SB volumes (cf. Fig 7). In 16 out of 19 patients BT detected non-enhancing tumor. The correlation matrix of the segmented tumor volumes is provided in Table 1. All computed correlations were rather high, with 0.8 being the lowest—for BT vs. GT. The semi-automatic SB method achieved a high inter-rater and GT correlation, although the latter value was overall slightly inferior.

## Sensitivity and PPV

BT achieved a high median sensitivity. The IQR was larger for BT than for SB and similar for all four SB raters (see Fig 5). The PPV of the SB raters showed good agreement, yet differed from BT. All raters and methods agreed on an even distribution of the PPV among patients. All four SB raters attained higher median PPV than sensitivity values, whereas BT appeared to be more consistent between PPV and sensitivity.

## Discussion

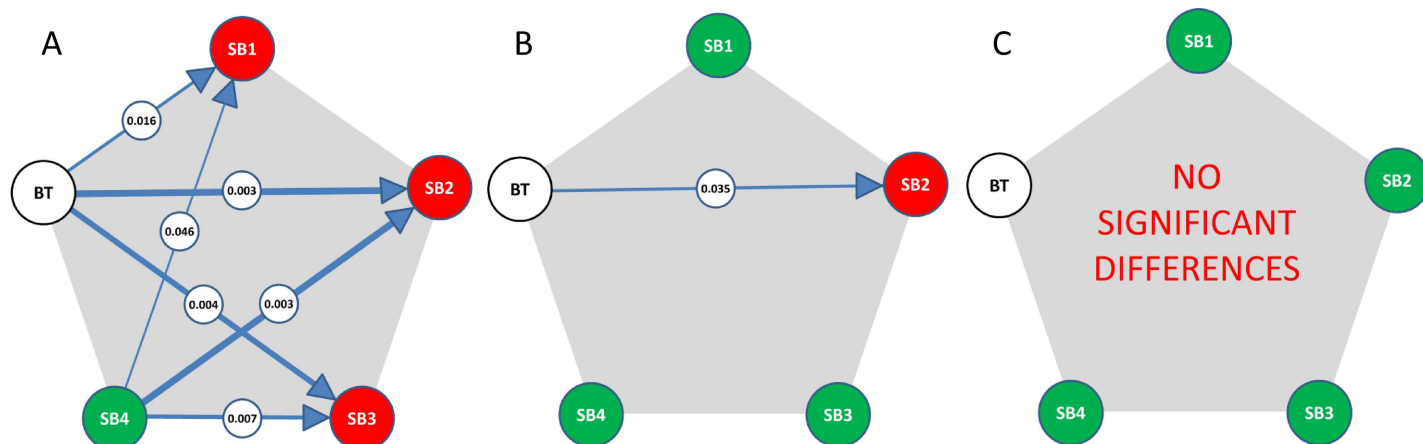
This study aimed to compare semi-automatic with fully-automatic brain tumor segmentation methods in terms of more clinically (SPD, volume) vs. technically relevant (Dice coefficient, PPV, sensitivity) metrics. While SB tended to be superior with respect to the metrics employed,



**Fig 3. Differences between the SPD metric and the GT for the BT and the four SB segmentations.** Additionally to the general boxplot statistics the mean value is shown (red asterisk). Negative values imply an overestimation by the rater/method whereas positive values indicate an underestimation.

doi:10.1371/journal.pone.0165302.g003





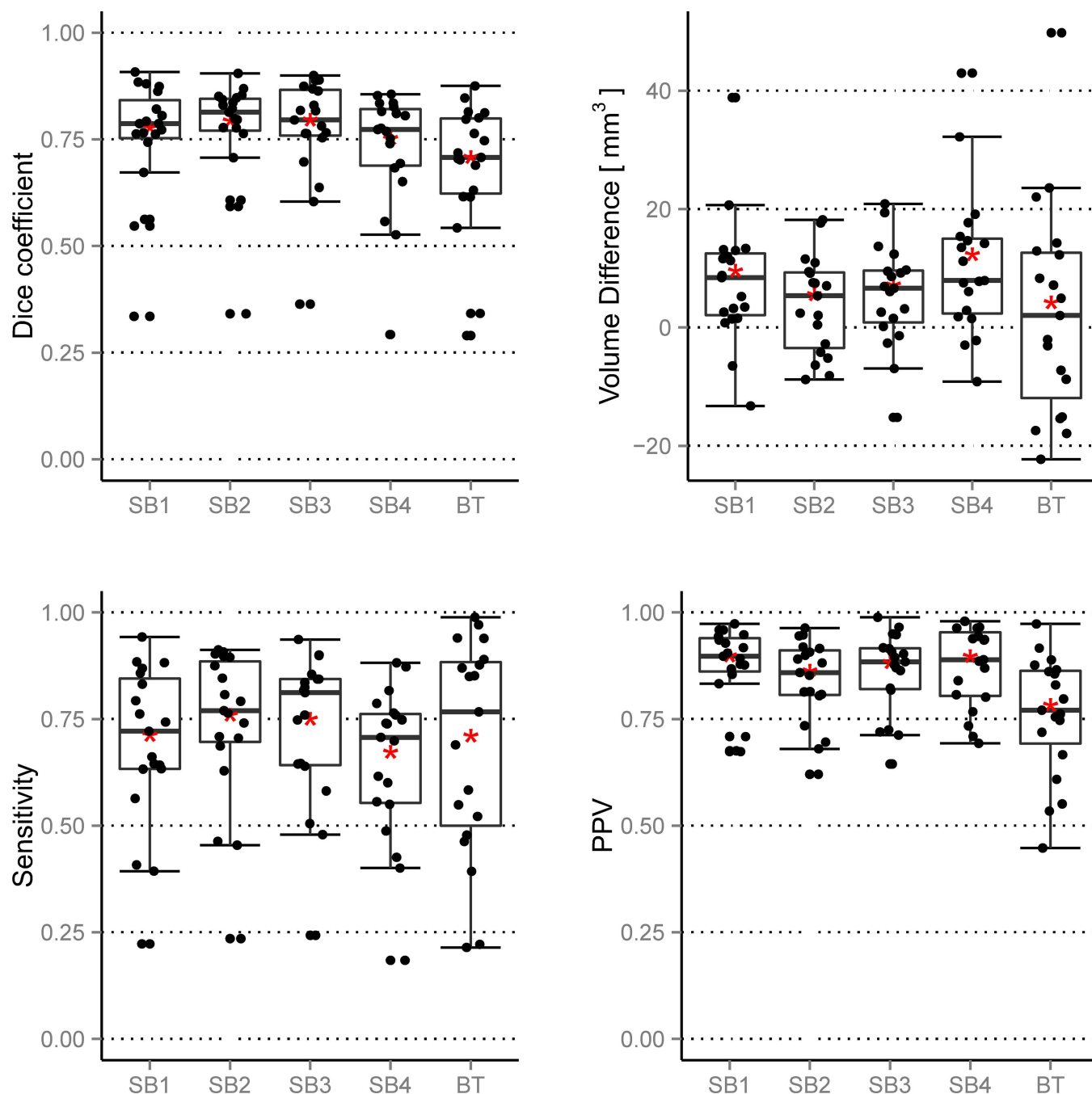
**Fig 4.** The performances of the different raters (SB1 to 4) and methods (SB vs. BT) were compared with a Bonferroni corrected Wilcoxon signed-rank test in terms of Dice coefficient (A), absolute volume difference (B) and absolute SPD difference (C). A connection denotes a significant difference, with the encircled number being the p-value. The arrow points to the superior method with respect to the GT. The line thickness depicts the p-value in a qualitative manner. The color coding shows to which SB rater BT (white) is significantly different (red) or not (green).

doi:10.1371/journal.pone.0165302.g004

we observed a discrepancy between technically and clinically applicable measures. The Dice coefficient, a standard metric in image analysis, reveals statistically significant differences not only between methods (BT vs. SB) but also between raters (four SB raters) of the same method. For the clinically relevant measures of the tumor volume, we observed only minor differences for one of the four raters, while the SPD calculations were in a comparable range.

SB can be reliably used for gross tumor segmentation as shown in [28]. However, inter-rater variability must be taken into account and is a principal drawback of user-dependent methods. Whether these variations primarily arise from selection bias of the perpendicular slices needed for interpolation, the variability of tumor delineation on these slices or the discrepancies in correction steps taken subsequently, is beyond the scope of this study. For BT, rater independence is an unarguable advantage in terms of segmentation consistency. This is of particular importance if tumor growth has to be addressed, e.g. in the framework of longitudinal clinical studies, pre-therapeutic tumor growth assessment before radiotherapy planning or for radiomics [29, 30]. Here, the subcompartment analysis of BT allows the integration of diffuse growth patterns beyond the disrupted blood–brain barrier that encompass the NETV. The automatic segmentation tool enabled visualization not only of the enhancing, but also of the T2/FLAIR-related GBM expansion within a single analysis. In this study, we observed extended NETV in 16/19 patients (confirmed by the GT), that were obscured by the SB segmentations. Intraoperative 5-aminolevulinic acid (5-ALA) fluorescence staining indicated that resection of non-enhancing tumor tissue may be useful in terms of overall survival, and that non-enhancing tumor compartments may be indicative of infiltrative and progressive tumor behavior [9, 11, 31, 32]. Non-enhancing tumor compartments were recently correlated with outcome and showed an impact of non-enhancing tumor on overall survival [16, 33]. As a consequence, the additional information provided by BT might further impact biopsy and resection planning, therapy selection [34–36] and monitoring intervals [37].

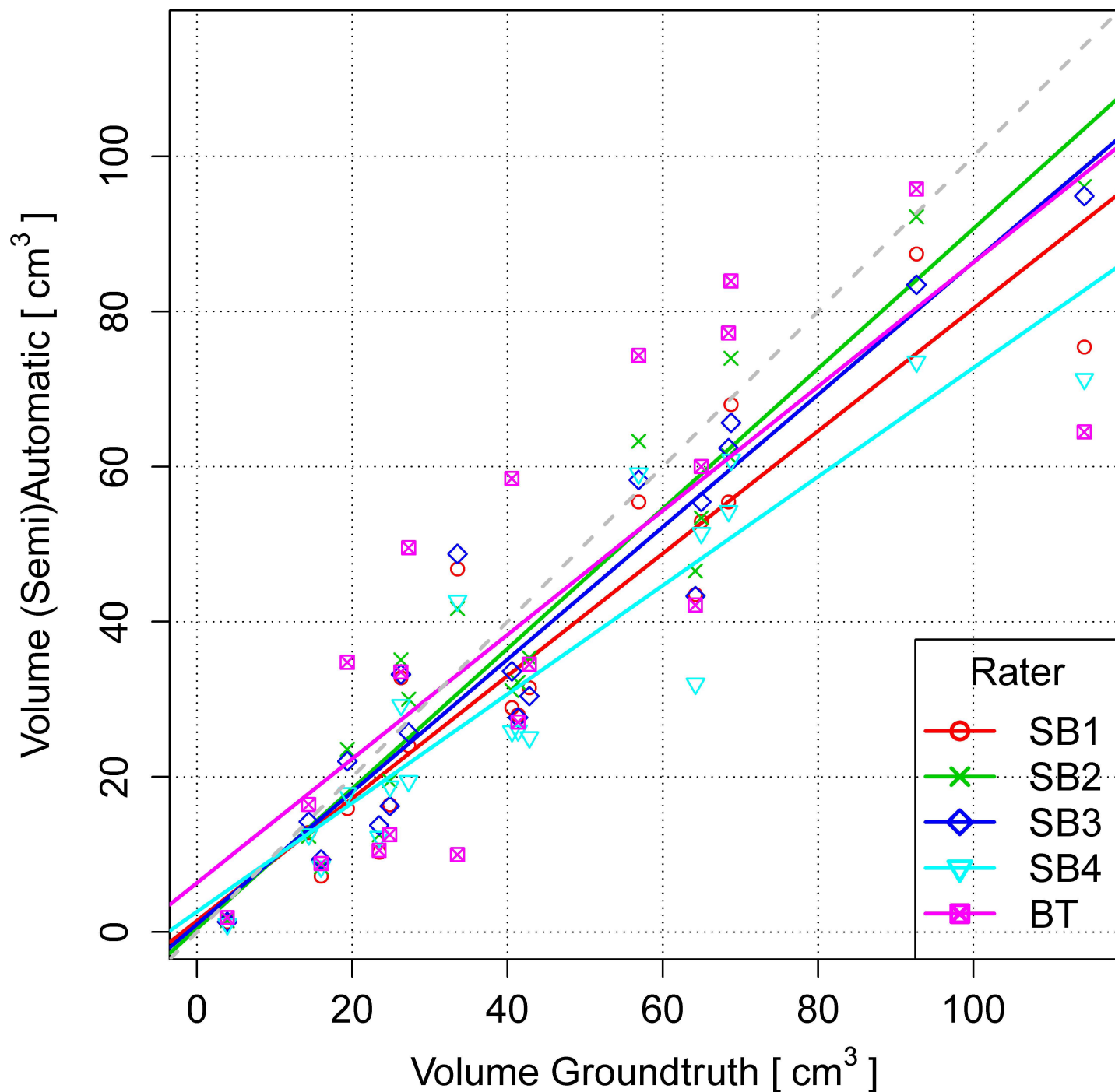
Tumors may consist of various heterogeneous tissue types that are shown by BT and that may improve knowledge of the heterogeneous cellular characteristics within the subcompartments of malignant gliomas beyond the areas of contrast uptake. More precise compartmental sampling may foster the development of strategies that target subtype-specific patterns of the tumor microenvironment.



**Fig 5. Left to right, top to bottom: Dice coefficient, volume difference, sensitivity and PPV.** The results are obtained when BT and the four SB raters are compared to the GT. The figure depicts the general boxplot statistics with the additional mean value (red asterisk). For the volume differences, negative values denote overestimation and positive values underestimation compared to the GT.

doi:10.1371/journal.pone.0165302.g005

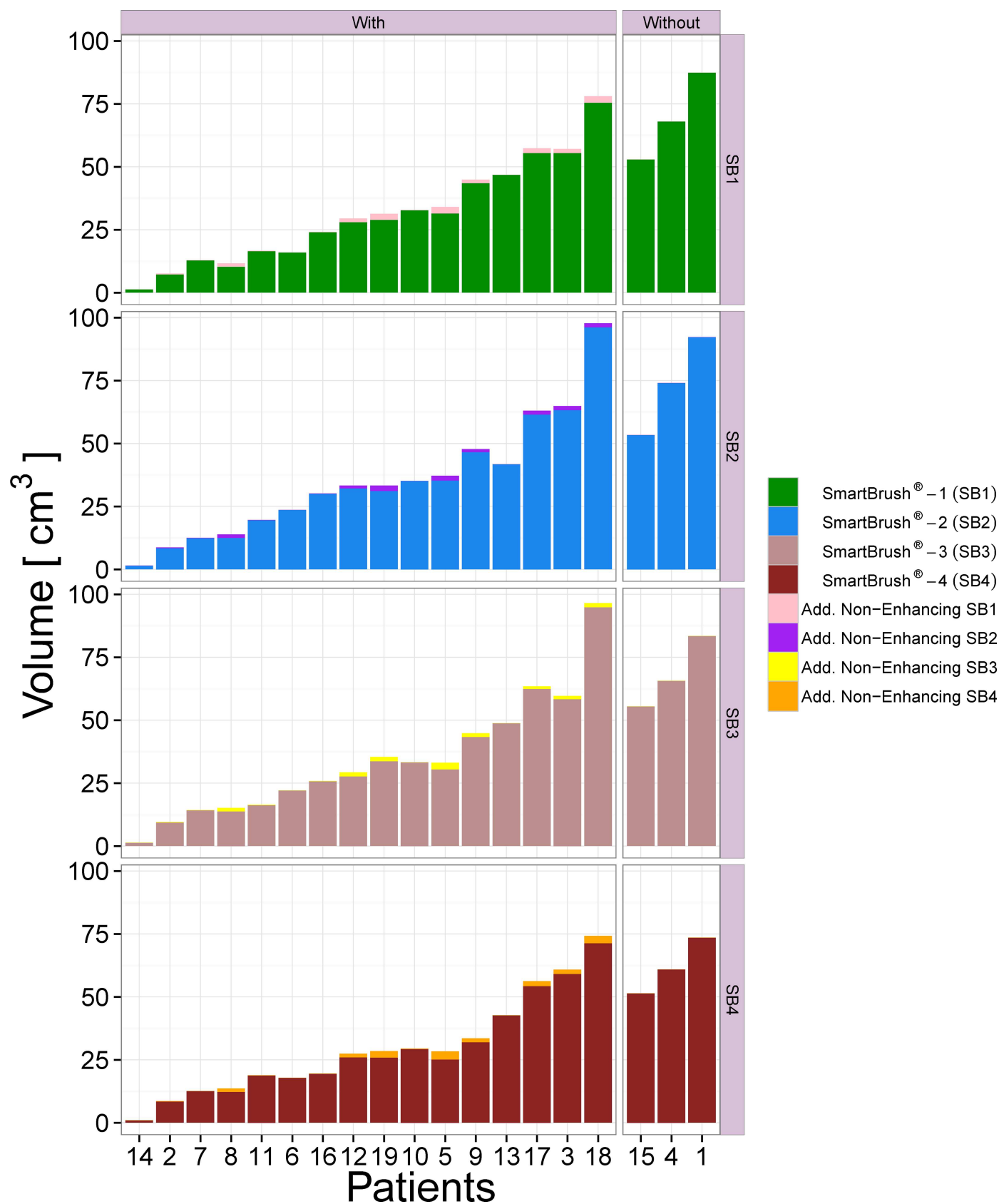
Processing times were surprisingly variable among the raters. Even with semi-automatic methods, each rater has to decide when he or she has achieved satisfactory segmentation results. One may accept the two minimally required perpendicular slices as satisfactory or continue with slice-by-slice corrections. These choices may result in a considerable time and quality difference, with the consequence of resembling either computer-assisted segmentation (as



**Fig 6. Volumes of the BT and the four SB segmentations (y-axis) plotted against the respective GT segmentations (x-axis).** Perfect agreement with respect to tumor volume means that all data points (volumes) would come to lie on the gray dashed 45-degree line starting from the origin (0,0).

doi:10.1371/journal.pone.0165302.g006

intended by the automatization procedures) or of bouncing back to human control (closer to manual segmentation). We did not control for this effect since we did not provide strict guidelines about the number of interactive steps to be taken to reach the final segmentation. However, this reflects good clinical practice and may in part explain the observed time divergences



**Fig 7. Barplot depicting the SB tumor volumes for the individual patients.** Non-enhancing tumor tissue that was, as confirmed by the GT, correctly segmented with the multi-compartment BT software and was not part of the SB segmented volume, is stacked on top of them. The barplot is horizontally split into two groups of patients with and without additional non-enhancing tissue found by BT. Vertically, the figure is quartered to show the results for all four SB raters.

doi:10.1371/journal.pone.0165302.g007

**Table 1. Correlation Matrix for Automatically (BT), Semi-Automatically (SB) and Manually Derived Tumor Volumes.<sup>a</sup>**

	SB1	SB2	SB3	SB4	BT	GT
SB1	1.00	0.99	0.98	0.99	0.85	0.93
SB2	0.99	1.00	0.99	0.98	0.86	0.96
SB3	0.98	0.99	1.00	0.98	0.83	0.96
SB4	0.99	0.98	0.98	1.00	0.85	0.92
BT	0.85	0.86	0.83	0.85	1.00	0.80
GT	0.93	0.96	0.96	0.92	0.80	1.00

<sup>a</sup>The Pearson correlation coefficient was employed.

doi:10.1371/journal.pone.0165302.t001

and inter-rater variability. Further, we recruited the raters according to their experience in glioma reading and their oncological expertise, which might have led to different strategies in preparing the final segmentation.

Finally, segmentation performance always reflects a mixture of additional non-professional skills such as self-motivation and ability to focus on the task, as well as the influence of task-induced fatigue and a multitude of other factors, including external ones. All these potential confounders can be overcome by automated segmentation.

## Conclusions

We conclude that automated tissue compartment analysis of GBMs using fully-automated analysis tools (BT) is feasible and provides similar results to those obtained with semi-automatic ones (SB) if the clinically relevant volume and SPD measures are primarily addressed. However the two methods differ considerably in terms of Dice coefficient due to rater-dependency. In addition, BT extends the GTV by adding NETV subcompartments omitted by a single-compartment (e.g. 3D T1w contrast-enhanced sequences) analysis, as is usually performed for therapy planning. Rater independence renders the tool applicable for complex data analysis (e.g. in radiosurgery planning), for tumor growth modeling and radiomic/radiogenomic analyses. Since both methods have complementary strengths (and limitations) their usage should be related to the clinical and scientific questions under consideration.

## Acknowledgments

We thank the MR technicians of our department for their excellent support. The authors also wish to acknowledge funding from Swiss National Foundation Grant number 140958.

## Author Contributions

**Conceptualization:** RW MR NP.

**Data curation:** RW MR NP SH.



**Formal analysis:** NP SH UK RV AJ JF.

**Funding acquisition:** PS JS MR RW.

**Investigation:** NP SH AJ JF UK.

**Methodology:** MR SH RW.

**Project administration:** NP.

**Resources:** MR RW AR.

**Software:** CR MR RM.

**Supervision:** MR RW.

**Validation:** PS JB AR UK.

**Visualization:** SH.

**Writing – original draft:** NP SH RW MR.

**Writing – review & editing:** NP SH RM RV AJ JF UK CR PS JB AR JS MR RW.

## References

1. Jiang J, Wu Y, Huang M, Yang W, Chen W, Feng Q. 3D brain tumor segmentation in multimodal MR images based on learning population- and patient-specific feature sets. *Computerized medical imaging and graphics: the official journal of the Computerized Medical Imaging Society*. [Research Support, Non-U.S. Gov't]. 2013 Oct-Dec; 37(7–8):512–21.
2. Mazzara GP, Velthuisen RP, Pearlman JL, Greenberg HM, Wagner H. Brain tumor target volume determination for radiation treatment planning through automated MRI segmentation. *International Journal of Radiation Oncology\* Biology\* Physics*. 2004; 59(1):300–12.
3. Deeley MA, Chen A, Datterli R, Noble JH, Cmelak AJ, Donnelly EF, et al. Comparison of manual and automatic segmentation methods for brain structures in the presence of space-occupying lesions: a multi-expert study. *Physics in medicine and biology*. 2011 Jul 21; 56(14):4557–77. doi: [10.1088/0031-9155/56/14/021](https://doi.org/10.1088/0031-9155/56/14/021) PMID: [21725140](https://pubmed.ncbi.nlm.nih.gov/21725140/)
4. Pica A, Terribilini D, Porz N, Reyes M, Slotboom J, Wiest R, et al. O2.03target Delineation In Glioblastoma: Is Preoperative Automatic Comparable To Expert Based Segmentation? *Neuro-Oncology*. 2014 September 1, 2014; 16(suppl 2):ii3.
5. Davatzikos C, Zacharaki EI, Gooya A, Clark V. Multi-parametric analysis and registration of brain tumors: constructing statistical atlases and diagnostic tools of predictive value. *Conference proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE Engineering in Medicine and Biology Society Annual Conference*. [Research Support, N.I.H., Extramural]. 2011;2011:6979–81.
6. Lacroix M, Abi-Said D, Fournay DR, Gokaslan ZL, Shi W, DeMonte F, et al. A multivariate analysis of 416 patients with glioblastoma multiforme: prognosis, extent of resection, and survival. *J Neurosurg*. 2001 Aug; 95(2):190–8. doi: [10.3171/jns.2001.95.2.0190](https://doi.org/10.3171/jns.2001.95.2.0190) PMID: [11780887](https://pubmed.ncbi.nlm.nih.gov/11780887/)
7. Stummer W, Kamp MA. The importance of surgical resection in malignant glioma. *Curr Opin Neurol*. 2009 Dec; 22(6):645–9. doi: [10.1097/WCO.0b013e3283320165](https://doi.org/10.1097/WCO.0b013e3283320165) PMID: [19738467](https://pubmed.ncbi.nlm.nih.gov/19738467/)
8. Stummer W, Reulen HJ, Meinel T, Pichlmeier U, Schumacher W, Tonn JC, et al. Extent of resection and survival in glioblastoma multiforme: identification of and adjustment for bias. *Neurosurgery*. 2008 Mar; 62(3):564–76; discussion -76. doi: [10.1227/01.neu.0000317304.31579.17](https://doi.org/10.1227/01.neu.0000317304.31579.17) PMID: [18425006](https://pubmed.ncbi.nlm.nih.gov/18425006/)
9. Schucht P, Knittel S, Slotboom J, Seidel K, Murek M, Jilch A, et al. 5-ALA complete resections go beyond MR contrast enhancement: shift corrected volumetric analysis of the extent of resection in surgery for glioblastoma. *Acta neurochirurgica*. 2014 Feb; 156(2):305–12; discussion 12. doi: [10.1007/s00701-013-1906-7](https://doi.org/10.1007/s00701-013-1906-7) PMID: [24449075](https://pubmed.ncbi.nlm.nih.gov/24449075/)
10. Coburger J, Hagel V, Wirtz CR, König R. Surgery for Glioblastoma: Impact of the Combined Use of 5-Aminolevulinic Acid and Intraoperative MRI on Extent of Resection and Survival. *PLoS One*. 2015; 10(6):e0131872. doi: [10.1371/journal.pone.0131872](https://doi.org/10.1371/journal.pone.0131872) PMID: [26115409](https://pubmed.ncbi.nlm.nih.gov/26115409/)
11. Coburger J, Engelke J, Scheuerle A, Thal DR, Hlavac M, Wirtz CR, et al. Tumor detection with 5-aminolevulinic acid fluorescence and Gd-DTPA-enhanced intraoperative MRI at the border of contrast-

- p>enhancing lesions: a prospective study based on histopathological assessment.
- Neurosurg Focus*
- . 2014 Feb; 36(2):E3. doi:
- [10.3171/2013.11.FOCUS13463](https://doi.org/10.3171/2013.11.FOCUS13463)
- PMID:
- [24484256](https://pubmed.ncbi.nlm.nih.gov/24484256/)
12. Sanai N, Polley MY, McDermott MW, Parsa AT, Berger MS. An extent of resection threshold for newly diagnosed glioblastomas. *J Neurosurg*. 2011 Jul; 115(1):3–8. doi: [10.3171/2011.2.JNS10998](https://doi.org/10.3171/2011.2.JNS10998) PMID: [21417701](https://pubmed.ncbi.nlm.nih.gov/21417701/)
  13. Menze B, Reyes M, Van Leemput K. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE transactions on medical imaging*. 2014 Dec 4.
  14. Bauer S1 NL, Reyes M. Fully automatic segmentation of brain tumor images using support vector machine classification in combination with hierarchical conditional random field regularization. *Med Image Comput Comput Assist Interv*. 2011; 14:354–61. PMID: [22003719](https://pubmed.ncbi.nlm.nih.gov/22003719/)
  15. Porz N, Bauer S, Pica A, Schuch P, Beck J, Verma RK, et al. Multi-modal glioblastoma segmentation: man versus machine. *PLoS one*. [Research Support, Non-U.S. Gov't]. 2014; 9(5):e96873. doi: [10.1371/journal.pone.0096873](https://doi.org/10.1371/journal.pone.0096873) PMID: [24804720](https://pubmed.ncbi.nlm.nih.gov/24804720/)
  16. Grabowski MM, Recinos PF, Nowacki AS, Schroeder JL, Angelov L, Barnett GH, et al. Residual tumor volume versus extent of resection: predictors of survival after surgery for glioblastoma. *Journal of neurosurgery*. [Research Support, Non-U.S. Gov't]. 2014 Nov; 121(5):1115–23. doi: [10.3171/2014.7.JNS132449](https://doi.org/10.3171/2014.7.JNS132449) PMID: [25192475](https://pubmed.ncbi.nlm.nih.gov/25192475/)
  17. Crum WR, Camara O, Hill DLG. Generalized overlap measures for evaluation and validation in medical image analysis. *IEEE transactions on medical imaging*. 2006; 25(11):1451–61. doi: [10.1109/TMI.2006.880587](https://doi.org/10.1109/TMI.2006.880587) PMID: [17117774](https://pubmed.ncbi.nlm.nih.gov/17117774/)
  18. Miller AB, Hoogstraten B, Staquet M, Winkler A. Reporting results of cancer treatment. *cancer*. 1981; 47(1):207–14. PMID: [7459811](https://pubmed.ncbi.nlm.nih.gov/7459811/)
  19. Gallego Perez-Larraya J, Lahutte M, Petirena G, Reyes-Botero G, Gonzalez-Aguilar A, Houillier C, et al. Response assessment in recurrent glioblastoma treated with irinotecan-bevacizumab: comparative analysis of the Macdonald, RECIST, RANO, and RECIST + F criteria. *Neuro Oncol*. 2012 May; 14(5):667–73. doi: [10.1093/neuonc/nos070](https://doi.org/10.1093/neuonc/nos070) PMID: [22492961](https://pubmed.ncbi.nlm.nih.gov/22492961/)
  20. Altman DG, Bland JM. Diagnostic tests. 1: Sensitivity and specificity. *BMJ: British Medical Journal*. 1994; 308(6943):1552. PMID: [8019315](https://pubmed.ncbi.nlm.nih.gov/8019315/)
  21. Altman DG, Bland JM. Statistics Notes: Diagnostic tests 2: predictive values. *Bmj*. 1994; 309(6947):102. PMID: [8038641](https://pubmed.ncbi.nlm.nih.gov/8038641/)
  22. Van Leemput K, Sabuncu MR, editors. A cautionary analysis of staple using direct inference of segmentation truth 2014: Springer.
  23. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin JC, Pujol S, et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magn Reson Imaging*. 2012 Nov; 30(9):1323–41. doi: [10.1016/j.mri.2012.05.001](https://doi.org/10.1016/j.mri.2012.05.001) PMID: [22770690](https://pubmed.ncbi.nlm.nih.gov/22770690/)
  24. Bauer S, Fejes T, Reyes M. A skull-stripping filter for ITK. *Insight Journal*. 2012.
  25. Bauer S, Fejes T, Slotboom J, Wiest R, Nolte L-P, Reyes M, editors. Segmentation of brain tumor images based on integrated hierarchical classification and regularization 2012.
  26. Wilcoxon F. Individual comparisons of grouped data by ranking methods. *Journal of economic entomology*. 1946; 39(2):269–70.
  27. Dunn OJ. Multiple comparisons among means. *Journal of the American Statistical Association*. 1961; 56(293):52–64.
  28. Huber T, Alber G, Bette S, Boeckh-Behrens T, Gempt J, Ringel F, et al. Reliability of Semi-Automated Segmentations in Glioblastoma. *Clinical neuroradiology*. 2015:1–9.
  29. Velazquez ER, Meier R, Dunn WD Jr, Alexander B, Wiest R, Bauer S, et al. Fully automatic GBM segmentation in the TCGA-GBM dataset: Prognosis and correlation with VASARI features. *Scientific reports*. 2015; 5.
  30. Rios VE, Meier R, Dunn W, Alexander B, Wiest R, Bauer S, et al. TU-AB-BRA-11: Evaluation of Fully Automatic Volumetric GBM Segmentation in the TCGA-GBM Dataset: Prognosis and Correlation with VASARI Features. *Medical physics*. 2015; 42(6):3589–.
  31. Aldave G, Tejada S, Pay E, Marigil M, Bejarano B, Idoate MA, et al. Prognostic value of residual fluorescent tissue in glioblastoma patients after gross total resection in 5-aminolevulinic Acid-guided surgery. *Neurosurgery*. 2013 Jun; 72(6):915–20; discussion 20–1. doi: [10.1227/NEU.0b013e31828c3974](https://doi.org/10.1227/NEU.0b013e31828c3974) PMID: [23685503](https://pubmed.ncbi.nlm.nih.gov/23685503/)
  32. Coburger J, Wirtz CR, König RW. Impact of extent of resection and recurrent surgery on clinical outcome and overall survival in a consecutive series of 170 patients for glioblastoma in intraoperative high field iMRI. *J Neurosurg Sci*. 2015 Jul 7.

33. Velazquez ER, Narayan V, Grossmann P, Dunn W, Gutman D, Aerts H. TU-CD-BRB-04: Automated Radiomic Features Complement the Prognostic Value of VASARI in the TCGA-GBM Dataset. *Med Phys*. 2015 Jun; 42(6):3603.
34. Schucht P, Beck J, Seidel K, Raabe A. Extending resection and preserving function: modern concepts of glioma surgery. *Swiss medical weekly*. 2015; 145:w14082. doi: [10.4414/smw.2015.14082](https://doi.org/10.4414/smw.2015.14082) PMID: [25651063](https://pubmed.ncbi.nlm.nih.gov/25651063/)
35. Schucht P, Seidel K, Beck J, Murek M, Jilch A, Wiest R, et al. Intraoperative monopolar mapping during 5-ALA-guided resections of glioblastomas adjacent to motor eloquent areas: evaluation of resection rates and neurological outcome. *Neurosurgical focus*. 2014 Dec; 37(6):E16. doi: [10.3171/2014.10.FOCUS14524](https://doi.org/10.3171/2014.10.FOCUS14524) PMID: [25434385](https://pubmed.ncbi.nlm.nih.gov/25434385/)
36. Raabe A, Beck J, Schucht P, Seidel K. Continuous dynamic mapping of the corticospinal tract during surgery of motor eloquent brain tumors: evaluation of a new method. *Journal of neurosurgery*. [Evaluation Studies]. 2014 May; 120(5):1015–24. doi: [10.3171/2014.1.JNS13909](https://doi.org/10.3171/2014.1.JNS13909) PMID: [24628613](https://pubmed.ncbi.nlm.nih.gov/24628613/)
37. Ellingson BM, Malkin MG, Rand SD, LaViolette PS, Connelly JM, Mueller WM, et al. Volumetric analysis of functional diffusion maps is a predictive imaging biomarker for cytotoxic and anti-angiogenic treatments in malignant gliomas. *J Neurooncol*. 2011 Mar; 102(1):95–103. doi: [10.1007/s11060-010-0293-7](https://doi.org/10.1007/s11060-010-0293-7) PMID: [20798977](https://pubmed.ncbi.nlm.nih.gov/20798977/)